

CLAIMS

- Sub B⁵
1. A method of searching a database to find documents similar to a query document, comprising:

decomposing the query document into elements of different data types;

10 for one or more of the elements in a first data type, conducting a first data type similarity search to return match results from the database for the one or more elements in the first data type;

15 for one or more of the elements in a second data type, conducting a second data type similarity search to return match results from the database for the one or more elements in the first data type;

combining the match results from the first data type similarity search and the second data type similarity search to provide query document match results.

- 20 2. A method as claimed in claim 1, wherein one of the data types is representative of text.

- 25 3. A method as claimed in claim 2, wherein a plurality of the data types are representative of text, separate data types of the plurality being representative of different functional blocks of text.

- Sub A¹
4. A method as claimed in any preceding claim, wherein one of the data types is representative of pictorial images.

- 30 5. A method as claimed in any preceding claim, wherein one of the data types is representative of graphical images.

6. A method as claimed in any preceding claim, wherein one of the data types is representative of the arrangement of other data types within the document.

7. A method as claimed in any preceding claim, wherein the step of similarity searching to return match results is carried out, separately, for a plurality of elements having between them more than two data types.
8. A method as claimed in any preceding claim, where all features of a common data type in the document are treated as one element.
9. A method as claimed in any of claims 1 to 7, where spatially distinct features of a common data type in the document are treated as separate elements.
10. A method as claimed in any preceding claim, wherein elements are user selectable or deselectable for the step of similarity searching.
11. A method as claimed in any preceding claim, wherein the similarity searching results for separate elements are weighted before combination.
12. A method as claimed in claim 11, wherein said weighting is user selected.
13. A method as claimed in claim 11, wherein said weighting is attributed according to a determined significance of each relevant element in the document.
14. A method of searching a database to find documents similar to a query document, comprising:
- decomposing the query document into elements of different data types;
- determining a layout element in a layout datatype from the spatial arrangement of the elements in the document;
- for the layout element, conducting a layout similarity search to return match results from the database for the layout element.

Sub A+

Sub B1
Don't

Sub A2

- ~~Sub-A3~~